

## Languages spoken by more people are more sound-symbolic

Shiri Lev-Ari, Ivet Kancheva, Louise Marston, Hannah Morris, Teah Swingler & Madina Zaynudinova (Royal Holloway, University of London, Egham, UK)

shiri.lev-ari@rhul.ac.uk

Languages are spoken in different social environments. These environments impose different communicative challenges. For example, larger communities might have less shared knowledge and greater difficulty of converging on a shared system. Recent research shows that languages adapt to their social environment [1], and in particular, that languages spoken by larger communities are less morphologically complex [1-2]. These findings suggest that languages spoken by more people might evolve to have features that make them more robust for learning and communication. This study tests this hypothesis by examining whether languages spoken by more people are more sound symbolic. This question is important not only because it sheds light on how linguistic features are shaped by communicative pressures, but also because it can address the question of the role of sound symbolism in language.

Sound symbolism has been shown to facilitate language learning and processing [e.g., 3-4]. Therefore, to the degree that languages spoken by more people should be more robust, they might rely more on sound symbolism. This hypothesis is in line with recent research on facial expressions that shows that more heterogeneous communities, which also face greater communicative challenges, use more exaggerated facial expressions, and these are better understood by non-community members [5].

To test whether languages spoken by more people are more sound symbolic, we selected 20 languages spoken by millions of people (Median=81.7m; range: 24.5m-1.1billion) and 20 languages spoken by only hundreds or thousands of people (Median=3,750; range: 200-314,000). Next, we generated recordings of the words 'large' and 'small' in those languages using a text-to-speech synthesizer. We selected the words 'large' and 'small' as there is an established link between high front vowels such as 'i' and 'e' and small size and low and back vowels such as 'a', 'o' and 'u' and large size [e.g., 6]. 128 participants heard the words in a random order, and for each word, they guessed whether it means 'small' or 'large'. If they were familiar with the word, they indicated that they knew the word and did not provide a guess.

A logistic mixed effects regression revealed that participants were better at guessing word meanings in languages spoken by many vs few people ( $\beta=-0.3$ ,  $SE=0.15$ ,  $z=-2$ ,  $p<0.05$ ; see Fig 1a). An exploratory analysis using the (log-transformed) number of speakers rather than a categorical predictor suggests greater influence of community size when communities are small ( $\beta=0.3$ ,  $SE=0.01$ ,  $z=2.1$ ,  $p<0.04$ ; see Fig 1b). Next, we examined whether participants relied on vowels to make their judgments. Participants exhibited the established sound-symbolic patterns: they were more likely to guess that a word means "large" the more back vs front vowels it had ( $\beta=0.16$ ,  $SE=0.05$ ,  $z=3.2$ ,  $p<0.01$ ; see Fig 2). Interestingly, widely-spoken languages were not more likely than less common languages to have front/back vowels to indicate small/large size ( $p>0.1$ ). This suggests that widely-spoken languages relied on other sound-symbolic cues.

This study shows that languages adapt to their social environment, and in particular, that widely-spoken languages are more sound symbolic than languages spoken by few people. We propose that this is driven by the need to overcome the greater communicative challenges involved in interaction in larger communities. Some research suggests that language lose their iconicity with time [e.g., 7 for sign languages]. This study suggests that having a larger community of speakers can lead to maintenance of this iconicity. Interestingly, results also showed that even though widely-spoken language were more sound symbolic, they were less likely to follow vowel-size congruency. It might therefore be the case that community size influences not only the degree to which languages are sound symbolic but the type of sound symbolism that languages develop.

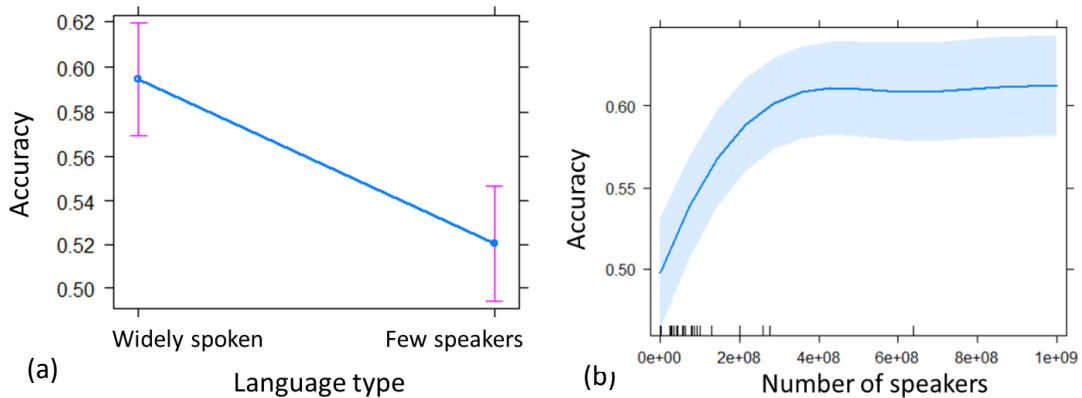


Figure 1

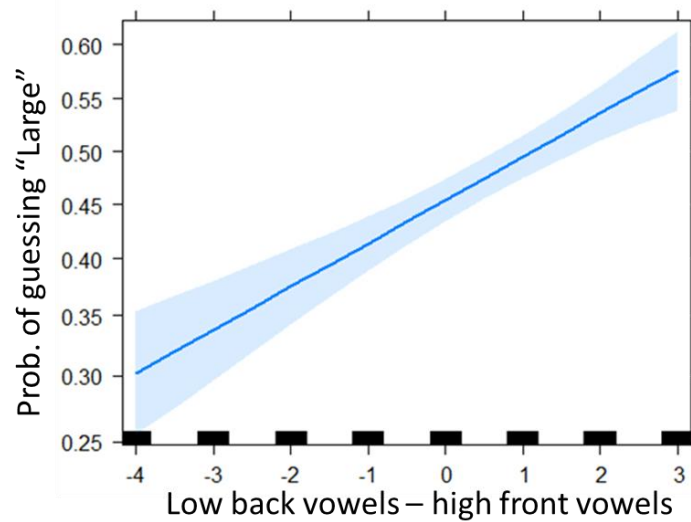


Figure 2

## References

- [1] Lupyán, G., & Dale, R. (2016). Why are there different languages? The role of adaptation in linguistic diversity. *Trends in cognitive sciences*, 20, 9, 649-660.
- [2] Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B*, 286, 1907, 20191262.
- [3] Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109, 54-65.
- [4] Meteyard, L., Stoppard, E., Snudden, D., Cappa, S. F., & Vigliocco, G. (2015). When semantics aids phonology: A processing advantage for iconic word forms in aphasia. *Neuropsychologia*, 76, 264-275.
- [5] Wood, A., Rychlowska, M., & Niedenthal, P. M. (2016). Heterogeneity of long-history migration predicts emotion recognition accuracy. *Emotion*, 16, 4, 413.
- [6] Peña, M., Mehler, J., & Nespór, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science*, 22, 1419-1421
- [7] Frishberg, N. (1975). Arbitrariness and iconicity: historical change in American Sign Language. *Language*, 696-719.